Proposal to encode the Hanifi Rohingya script in Unicode

Anshuman Pandey
Department of Linguistics
University of California, Berkeley
Berkeley, California, U.S.A.
anshuman.pandey@berkeley.edu

December 31, 2015

1 Introduction

This is a proposal to encode the Hanifi Rohingya script in Unicode. It supersedes the following document:

- L2/12-214 "Preliminary Proposal to Encode the Rohingya Script"

Apart from editoral improvements, the major changes from the preliminary proposal include:

- New names for several characters
- Redefinition of character properties
- Unification of the Rohingya *tatweel* with the corresponding Arabic character
- Inclusion of tables showing positional forms of letters

2 Background

The Hanifi Rohingya script (*ruwainggya*, *ruhainggya*) is used for writing Rohingya (ISO 639-3: rhg), an Indo-Aryan language spoken by one million people in Myanmar (Rakhine State) and by two-hundred thousand people in Bangladesh (Cox's Bazaar District). Rohingya is also spoken in several other countries as a result of the dispersion of the Rohingya community on account of persecution in Myanmar. The Rohingya language is closely related to Bengali. There are four different scripts used for writing the Rohingya language: Burmese, Arabic, the Latin-based 'Rohingylish', and the present script.

The script was developed by Maulana Mohammed Hanif and others in the 1980s. A chart showing the script as approved by a steering committee is shown in figure 1. It is modeled upon the Arabic script and shows the influence of other scripts. However, it is a constructed script and has no direct genetic affiliation to other scripts.

There is limited information available on the Hanifi Rohingya script in English. Most of the materials on the script are written in the Rohingya language. The Rohingya Language Committee and the Rohingya Education Board Myanmar have published multilingual primers of the script (see figures 3–16). There are also instructional videos available on YouTube (see figures 23–24). Educational materials and newspapers are published the script. Two typefaces have been developed by Muhammad Noor (see table 5).

3 Script Details

3.1 Structure

Hanifi Rohingya is an alphabetic script modeled upon Arabic. It is written right to left. Consonant letters possess an inherent vowel /ɔ/. The inherent vowel is changed by placing a vowel mark after the consonant. The inherent vowel is silenced by placing a special mark after a consonant letter. Three types of tone are indicating using combining signs placed above a letter or vowel mark. Nasalization is indicated using a mark placed after a letter or vowel mark. Consonant gemination is indicated using a sign placed above a letter.

The script is structurally conjoining. Letters join to following letters at the right edge at the baseline. In several hand-written sources, the conjunction between letters is broken at the baseline. This does not indicate any particular joining property of the unattached letter to the left of break. In texts printed using digitized fonts, the connections between letters at the baseline is consistently maintained. Positions forms of consonants and marks are given in tables 1 and 2.

3.2 Character names

Names for characters are based upon Latin transliterations for Rohignya characters given in various script primers (see tables 3 and 4). The names have been normalized in accordance with naming conventions for Indic scripts. This approach aligns with the Latin names given for the majority of characters across different publications. There are, however, different spellings for certain characters. For example, 𐴋 DDA is named '*rda*' in one source and *r̄a*' in another; 𐴈 KHA is alternately called '*xa*'; 𐴒 GA is also called '*gha*'. The 𐴚 NGA is known as *gan* and 𐴛 NYA is known as *nayya*. In cases where there are multiple names for a character, an appropriate name has been chosen based upon consultation with the user community. Names for tonal signs and other marks are based upon Indigenous names.

3.3 Representative glyphs

The representative glyphs are derived from the "Rohingya Gonya Leyka Noories" typeface developed by Muhammad Noor. The glyphs are used with permission. Original glyphs have been modified and additional glyphs have been introduced by the proposal author.

3.4 Letters

There are 28 letters:

	Character name	Joining	Value
𐴀	HANIFI ROHINGYA LETTER A	dual	/ɔ/, /ʔ/
𐴁	HANIFI ROHINGYA LETTER BA	dual	/b/
𐴂	HANIFI ROHINGYA LETTER PA	dual	/p/
𐴃	HANIFI ROHINGYA LETTER TA	dual	/t/
𐴄	HANIFI ROHINGYA LETTER TTA	dual	/ʈ/
𐴅	HANIFI ROHINGYA LETTER JA	dual	/ɟ/

𐴆	HANIFI ROHINGYA LETTER CA	dual	/c/
𐴇	HANIFI ROHINGYA LETTER HA	dual	/h/
𐴈	HANIFI ROHINGYA LETTER KHA	dual	/x/
𐴉	HANIFI ROHINGYA LETTER FA	dual	/f/
𐴊	HANIFI ROHINGYA LETTER DA	dual	/d/
𐴋	HANIFI ROHINGYA LETTER DDA	dual	/ɖ/
𐴌	HANIFI ROHINGYA LETTER RA	dual	/ɾ/
𐴍	HANIFI ROHINGYA LETTER RRA	dual	/ɽ/
𐴎	HANIFI ROHINGYA LETTER ZA	dual	/z/
𐴏	HANIFI ROHINGYA LETTER SA	dual	/s/
𐴐	HANIFI ROHINGYA LETTER SHA	dual	/ʃ/
𐴑	HANIFI ROHINGYA LETTER KA	dual	/k/
𐴒	HANIFI ROHINGYA LETTER GA	dual	/g/
𐴓	HANIFI ROHINGYA LETTER LA	dual	/l/
𐴔	HANIFI ROHINGYA LETTER MA	dual	/m/
𐴕	HANIFI ROHINGYA LETTER NA	dual	/n/
𐴖	HANIFI ROHINGYA LETTER WA	dual	/ʋ/, /v/
𐴗	HANIFI ROHINGYA LETTER SMALL WA	dual	/u/
𐴘	HANIFI ROHINGYA LETTER YA	dual	/j/
𐴙	HANIFI ROHINGYA LETTER SMALL YA	dual	/i/
𐴚	HANIFI ROHINGYA LETTER NGA	dual	/ŋ/
𐴛	HANIFI ROHINGYA LETTER NYA	dual	/ɲ/

The letter A is a vowel carrier. It represents the vowel /ɔ/. The letters SMALL WA (‘*kinna wa*’) and SMALL YA (‘*kinna ya*’) are used for representing /u/ and /i/ in diphthongs (see section 3.7).

The letter 𐴜 was previously used for representing /v/. Usage of this letter has been discontinued. The letter 𐴖 WA is now used in its place. It may be neccessary to encode 𐴜 (*ROHINGYA LETTER VA if it occurs in published material.

3.5 Vowel silencer

The following character is used for indicating a bare consonant:

	Character name	Joining	Value
𐴢	HANIFI ROHINGYA MARK SAKIN	right	—

The Rohingya *sakin* is similar in function to ْ U+0652 ARABIC SUKUN. It is placed after a letter in the encoded sequence, as shown below.

b	𐴁𐴢	<𐴁 BA, 𐴢 SAKIN>
p	𐴂𐴢	<𐴂 PA, 𐴢 SAKIN>
t	𐴃𐴢	<𐴃 TA, 𐴢 SAKIN>

The SAKIN is generally used with letters that occur in isolation or at the end of a word. There is no formal convention regarding its usage and it may be considered optional. For instance, the charts in figures 3, 4, and 5 show SAKIN written with certain consonants, yet the chart in figure 6 shows the sign written with every consonant letter. When SAKIN does occur, it is used generally with the following letters:

𐴁𐴢	𐴁	BA	𐴈𐴢	𐴈	KHA	𐴒𐴢	𐴒	GA
𐴂𐴢	𐴂	PA	𐴉𐴢	𐴉	FA	𐴓𐴢	𐴓	LA
𐴃𐴢	𐴃	TA	𐴊𐴢	𐴊	DA	𐴖𐴢	𐴖	WA
𐴄𐴢	𐴄	TTA	𐴏𐴢	𐴏	SA	𐴗𐴢	𐴗	SMALL WA
𐴇𐴢	𐴇	HA	𐴐𐴢	𐴐	SHA	𐴙𐴢	𐴙	SMALL YA

3.6 Final forms of letters

The following alternate forms are proposed for encoding:

	Character name	Joining	Value
	HANIFI ROHINGYA LETTER FINAL DA	right	/d/
	HANIFI ROHINGYA LETTER FINAL LA	right	/l/
	HANIFI ROHINGYA LETTER FINAL MA	right	/m/

These final forms are semantically and phonologically equivalent to the regular letter followed by *sakin*. They are used in word-final position or for representing the bare form of the respective consonant in isolated contexts.

FINAL DA	=	𐴊𐴢 <𐴊 DA, 𐴢 SAKIN>
FINAL LA	=	𐴓𐴢 <𐴓 LA, 𐴢 SAKIN>
FINAL MA	=	𐴔𐴢 <𐴔 MA, 𐴢 SAKIN>

The final letters may analyzed as stylistic representations of the above sequences. However, both representations occur in the available sources. They are proposed for inclusion in the repertoire in order to enable the complete encoding of Rohingya texts containing the final forms.

3.7 Vowel marks

There are 5 vowel marks:

	Character name	Joining	Value
𐴝	HANIFI ROHINGYA VOWEL MARK A	dual	/a/
𐴞	HANIFI ROHINGYA VOWEL MARK I	dual	/i/
𐴟	HANIFI ROHINGYA VOWEL MARK U	dual	/u/
𐴠	HANIFI ROHINGYA VOWEL MARK E	dual	/e/
𐴡	HANIFI ROHINGYA VOWEL MARK O	dual	/o/

When vowels occur in word-initial position, they are represented using a combination of the letter A and a vowel mark:

a	𐴀	<𐴀 A>
ā	𐴀𐴝	<𐴀 A, 𐴝 VOWEL MARK A>
i	𐴀𐴞	<𐴀 A, 𐴞 VOWEL MARK I>
u	𐴀𐴟	<𐴀 A, 𐴟 VOWEL MARK U>
e	𐴀𐴠	<𐴀 A, 𐴠 VOWEL MARK E>
o	𐴀𐴡	<𐴀 A, 𐴡 VOWEL MARK O>

When modifying a consonant, they are placed after the letter:

bā	𐴁𐴝	<𐴁 BA, 𐴝 VOWEL MARK A>
bi	𐴁𐴞	<𐴁 BA, 𐴞 VOWEL MARK I>
bu	𐴁𐴟	<𐴁 BA, 𐴟 VOWEL MARK U>
be	𐴁𐴠	<𐴁 BA, 𐴠 VOWEL MARK E>
bo	𐴁𐴡	<𐴁 BA, 𐴡 VOWEL MARK O>

Script charts show the vowel marks independently, unaccompanied by the letter A. These cases are intended for illustrating the mark and are not indicative of actual usage; the same applies to the other marks shown alongside the vowel marks in charts. Independent or initial vowels are expressed in combination with the letter A.

3.8 Diphthongs

Diphthongs are represented by placing 𐴙 SMALL YA and 𐴗 SMALL WA after a vowel mark. The SMALL YA represents the sound /i/ and SMALL WA represents /u/:

bai	𐴁𐴙	<𐴁 BA, 𐴙 SMALL YA>
bāi	𐴁𐴝𐴙	<𐴁 BA, 𐴝 VOWEL MARK A, 𐴙 SMALL YA>
bui	𐴁𐴟𐴙	<𐴁 BA, 𐴟 VOWEL MARK U, 𐴙 SMALL YA>
boi	𐴁𐴡𐴙	<𐴁 BA, 𐴡 VOWEL MARK O, 𐴙 SMALL YA>
bou	𐴁𐴡𐴗	<𐴁 BA, 𐴡 VOWEL MARK O, 𐴗 SMALL WA>

3.9 Nasalization mark

The following character is used for marking nasalization:

	Character name	Joining	Value
𐴣	HANIFI ROHINGYA MARK NA KHANNA	dual	/n/

It is placed after the consonant or vowel mark to be nasalized:

baṃ	𐴁𐴣	<𐴁 BA, 𐴣 NA KHANNA>
bāṃ	𐴁𐴝𐴣	<𐴁 BA, 𐴝 VOWEL MARK A, 𐴣 NA KHANNA>
baṃba	𐴁𐴣𐴁	<𐴁 BA, 𐴣 NA KHANNA, 𐴁 BA>

3.10 Tonal Signs

There are three tonal signs:

◌𐴤 HANIFI ROHINGYA SIGN HARBAHAY

◌𐴥 HANIFI ROHINGYA SIGN TAHALA

◌𐴦 HANIFI ROHINGYA SIGN TANA

Their properties are as follows:

		name	tone type	Arabic analogue
◌𐴤	HARBAHAY	*hārbahay*	short high tone	◌࣪ U+08EA ARABIC TONE ONE DOT ABOVE ◌࣭ U+08ED ARABIC TONE ONE DOT BELOW
◌𐴥	TAHALA	*ṭelā*	long falling tone	◌࣫ U+08EB ARABIC TONE TWO DOTS ABOVE ◌࣮ U+08EE ARABIC TONE TWO DOTS BELOW
◌𐴦	TANA	*ṭānā*	long rising tone	◌࣬ U+08EC ARABIC TONE LOOP ABOVE ◌࣯ U+08EF ARABIC TONE LOOP BELOW

The placement of tonal signs in the available sources is imprecise and not uniform. In some instances they are placed between the consonant and a vowel mark, in others they are positioned above the consonant, and in others upon the vowel; compare the placement of tones 2 and 3 in Figure 10, or compare lines 1 and and 2 and the examples in Figure 13.) These may be considered stylistic choices. In the encoded representation, a tone sign is placed after a consonant letter, but if a vowel mark is present, the tone sign may be placed either before or after the mark, depending upon orthography:

bá 𐴀𐴝𐴤 <𐴀 A, 𐴝 VOWEL MARK A, ◌𐴤 SIGN HARBAHAY>

bá 𐴀𐴤𐴝 <𐴀 A, ◌𐴤 SIGN HARBAHAY, 𐴝 VOWEL MARK A>

baa 𐴀𐴝𐴥 <𐴀 A, 𐴝 VOWEL MARK A, ◌𐴥 SIGN TAHALA>

baa 𐴀𐴥𐴝 <𐴀 A, ◌𐴥 SIGN TAHALA, 𐴝 VOWEL MARK A>

báa 𐴀𐴝𐴦 <𐴀 A, 𐴝 VOWEL MARK A, ◌𐴦 SIGN TANA>

báa 𐴀𐴦𐴝 <𐴀 A, ◌𐴦 SIGN TANA, 𐴝 VOWEL MARK A>

3.11 Gemination sign

The following character is used for indicating consonant gemination:

◌𐴧 HANIFI ROHINGYA SIGN TASSI

The Rohingya *tassi* is based upon and similar in function to ّ U+0651 ARABIC SHADDA. It is positioned above the letter that is doubled. The sign TASSI is placed after the respective letter in the encoded sequence:

baba	𐴁𐴁	<𐴁 BA, 𐴁 BA>
babba	𐴁𐴁𐴧	<𐴁 BA, 𐴁 BA, ◌𐴧 TASSI>

If there is a vowel mark after the doubled letter, then the TASSI is placed before the vowel mark in the encoded sequence:

babbā	𐴁𐴁𐴧𐴝	<𐴁 BA, 𐴁 BA, ◌𐴧 TASSI, 𐴝 VOWEL MARK A >

3.12 Digits

There is a full set of decimal digits:

𐴰	HANIFI ROHINGYA DIGIT ZERO	𐴵	HANIFI ROHINGYA DIGIT FIVE
𐴱	HANIFI ROHINGYA DIGIT ONE	𐴶	HANIFI ROHINGYA DIGIT SIX
𐴲	HANIFI ROHINGYA DIGIT TWO	𐴷	HANIFI ROHINGYA DIGIT SEVEN
𐴳	HANIFI ROHINGYA DIGIT THREE	𐴸	HANIFI ROHINGYA DIGIT EIGHT
𐴴	HANIFI ROHINGYA DIGIT FOUR	𐴹	HANIFI ROHINGYA DIGIT NINE

As in Arabic, the Rohingya digits are written from left to right. The Arabic style ٠ is attested as a glyphic variant for 𐴰 ZERO.

3.13 Punctuation

There is no script-specific punctuation. The U+002E FULL STOP is commonly used, as are Arabic signs, such as the ، U+060C ARABIC COMMA and ؟ U+061F ARABIC QUESTION MARK.

3.14 'Tatweel'

A *tatweel*-like character is used for graphical elongation or justification. The ـ U+0640 ARABIC TATWEEL should be used for representing *tatweel* in Rohingya.

3.15 Collation

The sort order for Hanifi Rohingya is as follows:

𐴀 A < 𐴝 VOWEL MARK A < 𐴞 VOWEL MARK I < 𐴟 VOWEL MARK U <
𐴠 VOWEL MARK E < 𐴡 VOWEL MARK O < 𐴁 BA < 𐴂 PA < 𐴃 TA < 𐴄 TTA < 𐴅 JA <
𐴆 CA < 𐴇 HA < 𐴈 KHA < 𐴉 FA < FINAL DA < 𐴊 DA < 𐴋 DDA < 𐴌 RA <

RRA < ZA < SHA < SA < KA < GA < FINAL LA < LA <

FINAL MA < MA < NA KHANNA < NA < WA < SMALL WA < YA <

SMALL YA < NGA < NYA < SAKIN

The letters FINAL DA, FINAL LA, FINAL MA are to be sorted equally to DA, LA, MA when followed by the mark SAKIN.

Further research is needed for determining if the collation order for SMALL WA and SMALL YA is primary or secondary.

4 Character Properties

4.1 Unicode Character Data

```
10D00;HANIFI ROHINGYA LETTER A;Lo;0;AL;;;;;N;;;;;
10D01;HANIFI ROHINGYA LETTER BA;Lo;0;AL;;;;;N;;;;;
10D02;HANIFI ROHINGYA LETTER PA;Lo;0;AL;;;;;N;;;;;
10D03;HANIFI ROHINGYA LETTER TA;Lo;0;AL;;;;;N;;;;;
10D04;HANIFI ROHINGYA LETTER TTA;Lo;0;AL;;;;;N;;;;;
10D05;HANIFI ROHINGYA LETTER JA;Lo;0;AL;;;;;N;;;;;
10D06;HANIFI ROHINGYA LETTER CA;Lo;0;AL;;;;;N;;;;;
10D07;HANIFI ROHINGYA LETTER HA;Lo;0;AL;;;;;N;;;;;
10D08;HANIFI ROHINGYA LETTER KHA;Lo;0;AL;;;;;N;;;;;
10D09;HANIFI ROHINGYA LETTER FA;Lo;0;AL;;;;;N;;;;;
10D0A;HANIFI ROHINGYA LETTER DA;Lo;0;AL;;;;;N;;;;;
10D0B;HANIFI ROHINGYA LETTER DDA;Lo;0;AL;;;;;N;;;;;
10D0C;HANIFI ROHINGYA LETTER RA;Lo;0;AL;;;;;N;;;;;
10D0D;HANIFI ROHINGYA LETTER RRA;Lo;0;AL;;;;;N;;;;;
10D0E;HANIFI ROHINGYA LETTER ZA;Lo;0;AL;;;;;N;;;;;
10D0F;HANIFI ROHINGYA LETTER SA;Lo;0;AL;;;;;N;;;;;
10D10;HANIFI ROHINGYA LETTER SHA;Lo;0;AL;;;;;N;;;;;
10D11;HANIFI ROHINGYA LETTER KA;Lo;0;AL;;;;;N;;;;;
10D12;HANIFI ROHINGYA LETTER GA;Lo;0;AL;;;;;N;;;;;
10D13;HANIFI ROHINGYA LETTER LA;Lo;0;AL;;;;;N;;;;;
10D14;HANIFI ROHINGYA LETTER MA;Lo;0;AL;;;;;N;;;;;
10D15;HANIFI ROHINGYA LETTER NA;Lo;0;AL;;;;;N;;;;;
10D16;HANIFI ROHINGYA LETTER WA;Lo;0;AL;;;;;N;;;;;
10D17;HANIFI ROHINGYA LETTER SMALL WA;Lo;0;AL;;;;;N;;;;;
10D18;HANIFI ROHINGYA LETTER YA;Lo;0;AL;;;;;N;;;;;
10D19;HANIFI ROHINGYA LETTER SMALL YA;Lo;0;AL;;;;;N;;;;;
10D1A;HANIFI ROHINGYA LETTER NGA;Lo;0;AL;;;;;N;;;;;
10D1B;HANIFI ROHINGYA LETTER NYA;Lo;0;AL;;;;;N;;;;;
10D1C;HANIFI ROHINGYA MARK SUKUN;Lo;0;AL;;;;;N;;;;;
10D1D;HANIFI ROHINGYA VOWEL MARK A;Lo;0;AL;;;;;N;;;;;
10D1E;HANIFI ROHINGYA VOWEL MARK I;Lo;0;AL;;;;;N;;;;;
10D1F;HANIFI ROHINGYA VOWEL MARK U;Lo;0;AL;;;;;N;;;;;
10D20;HANIFI ROHINGYA VOWEL MARK E;Lo;0;AL;;;;;N;;;;;
10D21;HANIFI ROHINGYA VOWEL MARK O;Lo;0;AL;;;;;N;;;;;
10D22;HANIFI ROHINGYA MARK NA KHANNA;Lo;0;AL;;;;;N;;;;;
10D23;HANIFI ROHINGYA SIGN HARBAHAY;Mn;230;NSM;;;;;N;;;;;
10D24;HANIFI ROHINGYA SIGN TAHALA;Mn;230;NSM;;;;;N;;;;;
10D25;HANIFI ROHINGYA SIGN TANA;Mn;230;NSM;;;;;N;;;;;
10D26;HANIFI ROHINGYA SIGN SHADDA;Mn;230;NSM;;;;;N;;;;;
10D27;HANIFI ROHINGYA LETTER FINAL DA;Lo;0;AL;;;;;N;;;;;
10D28;HANIFI ROHINGYA LETTER FINAL LA;Lo;0;AL;;;;;N;;;;;
```

```
10D29;HANIFI ROHINGYA LETTER FINAL MA;Lo;0;AL;;;;;N;;;;;
10D30;HANIFI ROHINGYA DIGIT ZERO;Nd;0;AN;;0;0;0;N;;;;;
10D31;HANIFI ROHINGYA DIGIT ONE;Nd;0;AN;;1;1;1;N;;;;;
10D32;HANIFI ROHINGYA DIGIT TWO;Nd;0;AN;;2;2;2;N;;;;;
10D33;HANIFI ROHINGYA DIGIT THREE;Nd;0;AN;;3;3;3;N;;;;;
10D34;HANIFI ROHINGYA DIGIT FOUR;Nd;0;AN;;4;4;4;N;;;;;
10D35;HANIFI ROHINGYA DIGIT FIVE;Nd;0;AN;;5;5;5;N;;;;;
10D36;HANIFI ROHINGYA DIGIT SIX;Nd;0;AN;;6;6;6;N;;;;;
10D37;HANIFI ROHINGYA DIGIT SEVEN;Nd;0;AN;;7;7;7;N;;;;;
10D38;HANIFI ROHINGYA DIGIT EIGHT;Nd;0;AN;;8;8;8;N;;;;;
10D39;HANIFI ROHINGYA DIGIT NINE;Nd;0;AN;;9;9;9;N;;;;;
```

4.2 Arabic Shaping Data

```
10D00; HANIFI ROHINGYA LETTER A; D; A
10D01; HANIFI ROHINGYA LETTER BA; D; BA
10D02; HANIFI ROHINGYA LETTER PA; D; PA
10D03; HANIFI ROHINGYA LETTER TA; D; TA
10D04; HANIFI ROHINGYA LETTER TTA; D; TTA
10D05; HANIFI ROHINGYA LETTER JA; D; JA
10D06; HANIFI ROHINGYA LETTER CA; D; CA
10D07; HANIFI ROHINGYA LETTER HA; D; HA
10D08; HANIFI ROHINGYA LETTER KHA; D; KHA
10D09; HANIFI ROHINGYA LETTER FA; D; PA
10D0A; HANIFI ROHINGYA LETTER DA; D; DA
10D0B; HANIFI ROHINGYA LETTER DDA; D; DDA
10D0C; HANIFI ROHINGYA LETTER RA; D; RA
10D0D; HANIFI ROHINGYA LETTER RRA; D; RRA
10D0E; HANIFI ROHINGYA LETTER ZA; D; ZA
10D0F; HANIFI ROHINGYA LETTER SA; D; SA
10D10; HANIFI ROHINGYA LETTER SHA; D; SHA
10D11; HANIFI ROHINGYA LETTER KA; D; KA
10D12; HANIFI ROHINGYA LETTER GA; D; GA
10D13; HANIFI ROHINGYA LETTER LA; D; LA
10D14; HANIFI ROHINGYA LETTER MA; D; MA
10D15; HANIFI ROHINGYA LETTER NA; D; NA
10D16; HANIFI ROHINGYA LETTER WA; D; WA
10D17; HANIFI ROHINGYA LETTER SMALL WA; SMALL WA
10D18; HANIFI ROHINGYA LETTER YA; D; YA
10D19; HANIFI ROHINGYA LETTER SMALL YA; D; SAKIN
10D1A; HANIFI ROHINGYA LETTER NGA; D; NGA
10D1B; HANIFI ROHINGYA LETTER NYA; D; NYA
10D1C; HANIFI ROHINGYA MARK SAKIN; R; MARK SAKIN
10D1D; HANIFI ROHINGYA VOWEL MARK A; D; VOWEL MARK A
10D1E; HANIFI ROHINGYA VOWEL MARK I; D; NA KHANNA
10D1F; HANIFI ROHINGYA VOWEL MARK U; D; VOWEL MARK U
10D20; HANIFI ROHINGYA VOWEL MARK E; D; VOWEL MARK E
10D21; HANIFI ROHINGYA VOWEL MARK O; D; VOWEL MARK O
10D22; HANIFI ROHINGYA MARK NA KHANNA; D; NA KHANNA
10D27; HANIFI ROHINGYA LETTER FINAL DA; R; FINAL DA
10D28; HANIFI ROHINGYA LETTER FINAL LA; R; FINAL LA
10D29; HANIFI ROHINGYA LETTER FINAL MA; R; FINAL MA
```

4.3 Script Extensions

The following characters should be extended for usage in Rohingya:

```
0640 ; # Lm  ARABIC TATWEEL
```

```
060C ; # Po  ARABIC COMMA
061B ; # Po  ARABIC SEMICOLON
061F ; # Po  ARABIC QUESTION MARK
```

4.4 'Confusable' Characters

Some Rohingya letters resemble those found of Arabic. Attention to these 'confusables' is particularly important because Arabic is commonly used in Rohingya documents.

```
10D03 HANIFI ROHINGYA LETTER PA         ;  0648 ARABIC LETTER WAW
10D07 HANIFI ROHINGYA LETTER HA         ;  06BE ARABIC LETTER HEH DOACHASHMEE
10D08 HANIFI ROHINGYA LETTER KHA        ;  06A9 ARABIC LETTER KEHEH
10D09 HANIFI ROHINGYA LETTER FA         ;  06CF ARABIC LETTER WAW WITH DOT ABOVE
10D0B HANIFI ROHINGYA LETTER DDA        ;  0637 ARABIC LETTER TAH
10D0C HANIFI ROHINGYA LETTER RRA        ;  0637 ARABIC LETTER TAH
10D11 HANIFI ROHINGYA LETTER KA         ;  0637 ARABIC LETTER TAH
10D12 HANIFI ROHINGYA LETTER GA         ;  FECB ARABIC LETTER AIN INITIAL FORM
10D13 HANIFI ROHINGYA LETTER LA         ;  0644 ARABIC LETTER LAM
10D14 HANIFI ROHINGYA LETTER MA         ;  FEE3 ARABIC LETTER MEEM INITIAL FORM
10D19 HANIFI ROHINGYA LETTER SMALL YA   ;  FE91 ARABIC LETTER BEH INITIAL FORM
10D1D HANIFI ROHINGYA LETTER NA KHANNA  ;  0632 ARABIC LETTER ZAIN
10D1D HANIFI ROHINGYA LETTER NA KHANNA  ;  FEE7 ARABIC LETTER NOON INITIAL FORM
10D1E HANIFI ROHINGYA VOWEL MARK A      ;  FEBB ARABIC LETTER SAD INITIAL FORM
10D24 HANIFI ROHINGYA SIGN TONE-2       ;  0654 ARABIC HAMZA ABOVE
10D25 HANIFI ROHINGYA SIGN TONE-3       ;  0653 ARABIC MADDAH ABOVE
10D26 HANIFI ROHINGYA SIGN SHADDA       ;  0651 ARABIC SHADDA
```

5 References

noorismail52. 2011a. "Rohingya mother Language (letters) Part 1.flv". http://www.youtube.com/watch?v=w4h6w6NyvOU

———. 2011b. "Rohingya mother Language (letters) Part 3.flv". http://www.youtube.com/watch?v=pylvjjTQG8c

Pandey, Anshuman. 2012. "Preliminary Proposal to Encode the Rohingya Script" (L2/12-214). June 20, 2012. http://www.unicode.org/L2/L2012/12214-rohingya.pdf

Rohingya Education Board Myanmar. *Ruhainggya Zubanor Fonna: Hisab* ["Script of the Rohingya language: Counting"]. Ek kelasottu dui kelas: lego ar foro ["From Class 1 to Class 2: Read and Write"].

Rohingya Language Committee. [A]. *Kayda Ruwainggya Zubanor* [Primer of the Rohingya Language].

———. [B]. *Ruwaingya Zubanor Foyla Kitab* [First Book of the Rohingya Language].

6 Acknowledgments

I am thankful to Mattias Persson and Ian James for bringing the Rohingya script to my attention. Lorna Priest introduced me to James Lloyd-Williams, who very generously provided copies of Rohingya primers. I am all the more grateful to Muhammad Noor, who provided me with numerous books and other materials written and printed in the script. This proposal would not be possible without the assistance of these individuals.

This project was made possible in part through a Google Research Award, granted to Deborah Anderson for the Script Encoding Initiative, and a grant from the United States National Endowment for the Humanities (PR-50205-15), which funds the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of Google or the National Endowment for the Humanities.

	10D0	10D1	10D2	10D3
0	10D00	10D10	10D20	10D30
1	10D01	10D11	10D21	10D31
2	10D02	10D12	10D22	10D32
3	10D03	10D13	10D23	10D33
4	10D04	10D14	10D24	10D34
5	10D05	10D15	10D25	10D35
6	10D06	10D16	10D26	10D36
7	10D07	10D17	10D27	10D37
8	10D08	10D18	10D28	10D38
9	10D09	10D19	10D29	10D39
A	10D0A	10D1A		
B	10D0B	10D1B		
C	10D0C	10D1C		
D	10D0D	10D1D		
E	10D0E	10D1E		
F	10D0F	10D1F		

Letters

10D00 𐴀 HANIFI ROHINGYA LETTER A
10D01 𐴁 HANIFI ROHINGYA LETTER BA
10D02 𐴂 HANIFI ROHINGYA LETTER PA
10D03 𐴃 HANIFI ROHINGYA LETTER TA
10D04 𐴄 HANIFI ROHINGYA LETTER TTA
10D05 𐴅 HANIFI ROHINGYA LETTER JA
10D06 𐴆 HANIFI ROHINGYA LETTER CA
10D07 𐴇 HANIFI ROHINGYA LETTER HA
10D08 𐴈 HANIFI ROHINGYA LETTER KHA
10D09 𐴉 HANIFI ROHINGYA LETTER FA
10D0A 𐴊 HANIFI ROHINGYA LETTER DA
10D0B 𐴋 HANIFI ROHINGYA LETTER DDA
10D0C 𐴌 HANIFI ROHINGYA LETTER RA
10D0D 𐴍 HANIFI ROHINGYA LETTER RRA
10D0E 𐴎 HANIFI ROHINGYA LETTER ZA
10D0F 𐴏 HANIFI ROHINGYA LETTER SA
10D10 𐴐 HANIFI ROHINGYA LETTER SHA
10D11 𐴑 HANIFI ROHINGYA LETTER KA
10D12 𐴒 HANIFI ROHINGYA LETTER GA
10D13 𐴓 HANIFI ROHINGYA LETTER LA
10D14 𐴔 HANIFI ROHINGYA LETTER MA
10D15 𐴕 HANIFI ROHINGYA LETTER NA
10D16 𐴖 HANIFI ROHINGYA LETTER WA
10D17 𐴗 HANIFI ROHINGYA LETTER SMALL WA
= kinna wa
10D18 𐴘 HANIFI ROHINGYA LETTER YA
10D19 𐴙 HANIFI ROHINGYA LETTER SMALL YA
= kinna ya
10D1A 𐴚 HANIFI ROHINGYA LETTER NGA
= gan
10D1B 𐴛 HANIFI ROHINGYA LETTER NYA
= nayya

Vowel silencer

10D1C HANIFI ROHINGYA SIGN SAKIN

Vowel marks

10D1D HANIFI ROHINGYA VOWEL MARK A
= aa-for
10D1E HANIFI ROHINGYA VOWEL MARK I
= i-for
10D1F HANIFI ROHINGYA VOWEL MARK U
= u-for
10D20 HANIFI ROHINGYA VOWEL MARK E
= e-for
10D21 HANIFI ROHINGYA VOWEL MARK O
= o-for

Nasalization mark

10D22 HANIFI ROHINGYA MARK NA KHANNA

Tone marks

10D23 HANIFI ROHINGYA SIGN HARBAHAY
• short high tone
10D24 HANIFI ROHINGYA SIGN TAHALA
• long falling tone
10D25 HANIFI ROHINGYA SIGN TANA
• long rising tone

Gemination sign

10D26 HANIFI ROHINGYA SIGN TASSI

Final letters

10D27 HANIFI ROHINGYA FINAL LETTER DA
10D28 HANIFI ROHINGYA FINAL LETTER LA
10D29 HANIFI ROHINGYA FINAL LETTER MA

Digits

10D30 𐴰 HANIFI ROHINGYA DIGIT ZERO
10D31 𐴱 HANIFI ROHINGYA DIGIT ONE
10D32 𐴲 HANIFI ROHINGYA DIGIT TWO
10D33 𐴳 HANIFI ROHINGYA DIGIT THREE
10D34 𐴴 HANIFI ROHINGYA DIGIT FOUR
10D35 𐴵 HANIFI ROHINGYA DIGIT FIVE
10D36 𐴶 HANIFI ROHINGYA DIGIT SIX
10D37 𐴷 HANIFI ROHINGYA DIGIT SEVEN
10D38 𐴸 HANIFI ROHINGYA DIGIT EIGHT
10D39 𐴹 HANIFI ROHINGYA DIGIT NINE

	Character name	Joining	Final	Medial	Initial
𐴀	HANIFI ROHINGYA LETTER A	dual	‍𐴀	‍𐴀‍	𐴀‍
𐴁	HANIFI ROHINGYA LETTER BA	dual	‍𐴁	‍𐴁‍	𐴁‍
𐴂	HANIFI ROHINGYA LETTER PA	dual	‍𐴂	‍𐴂‍	𐴂‍
𐴃	HANIFI ROHINGYA LETTER TA	dual	‍𐴃	‍𐴃‍	𐴃‍
𐴄	HANIFI ROHINGYA LETTER TTA	dual	‍𐴄	‍𐴄‍	𐴄‍
𐴅	HANIFI ROHINGYA LETTER JA	dual	‍𐴅	‍𐴅‍	𐴅‍
𐴆	HANIFI ROHINGYA LETTER CA	dual	‍𐴆	‍𐴆‍	𐴆‍
𐴇	HANIFI ROHINGYA LETTER HA	dual	‍𐴇	‍𐴇‍	𐴇‍
𐴈	HANIFI ROHINGYA LETTER KHA	dual	‍𐴈	‍𐴈‍	𐴈‍
𐴉	HANIFI ROHINGYA LETTER FA	dual	‍𐴉	‍𐴉‍	𐴉‍
𐴊	HANIFI ROHINGYA LETTER DA	dual	‍𐴊	‍𐴊‍	𐴊‍
𐴋	HANIFI ROHINGYA LETTER DDA	dual	‍𐴋	‍𐴋‍	𐴋‍
𐴌	HANIFI ROHINGYA LETTER RA	dual	‍𐴌	‍𐴌‍	𐴌‍
𐴍	HANIFI ROHINGYA LETTER RRA	dual	‍𐴍	‍𐴍‍	𐴍‍
𐴎	HANIFI ROHINGYA LETTER ZA	dual	‍𐴎	‍𐴎‍	𐴎‍
𐴏	HANIFI ROHINGYA LETTER SA	dual	‍𐴏	‍𐴏‍	𐴏‍
𐴐	HANIFI ROHINGYA LETTER SHA	dual	‍𐴐	‍𐴐‍	𐴐‍
𐴑	HANIFI ROHINGYA LETTER KA	dual	‍𐴑	‍𐴑‍	𐴑‍
𐴒	HANIFI ROHINGYA LETTER GA	dual	‍𐴒	‍𐴒‍	𐴒‍
𐴓	HANIFI ROHINGYA LETTER LA	dual	‍𐴓	‍𐴓‍	𐴓‍
𐴔	HANIFI ROHINGYA LETTER MA	dual	‍𐴔	‍𐴔‍	𐴔‍
𐴕	HANIFI ROHINGYA LETTER NA	dual	‍𐴕	‍𐴕‍	𐴕‍
𐴖	HANIFI ROHINGYA LETTER WA	dual	‍𐴖	‍𐴖‍	𐴖‍
𐴗	HANIFI ROHINGYA LETTER SMALL WA	dual	‍𐴗	‍𐴗‍	𐴗‍
𐴘	HANIFI ROHINGYA LETTER YA	dual	‍𐴘	‍𐴘‍	𐴘‍
𐴙	HANIFI ROHINGYA LETTER SMALL YA	dual	‍𐴙	‍𐴙‍	𐴙‍
𐴚	HANIFI ROHINGYA LETTER NGA	dual	‍𐴚	‍𐴚‍	𐴚‍
𐴛	HANIFI ROHINGYA LETTER NYA	dual	‍𐴛	‍𐴛‍	𐴛‍

Table 1: Positional forms of consonant letters.

	Character name	Joining	Final	Medial	Initial
𐴝	HANIFI ROHINGYA VOWEL MARK A	dual	‍𐴝	‍𐴝‍	𐴝‍
𐴞	HANIFI ROHINGYA VOWEL MARK I	dual	‍𐴞	‍𐴞‍	𐴞‍
𐴟	HANIFI ROHINGYA VOWEL MARK U	dual	‍𐴟	‍𐴟‍	𐴟‍
𐴠	HANIFI ROHINGYA VOWEL MARK E	dual	‍𐴠	‍𐴠‍	𐴠‍
𐴡	HANIFI ROHINGYA VOWEL MARK O	dual	‍𐴡	‍𐴡‍	𐴡‍
𐴢	HANIFI ROHINGYA MARK SAKIN	right	‍𐴢	‍𐴢‍	𐴢
𐴣	HANIFI ROHINGYA MARK NA KHANNA	dual	‍𐴣	‍𐴣‍	𐴣‍

Table 2: Positional forms of various marks.

	Proposed character name	Fig. 4	Fig. 5	Fig. 6
𐴀	HANIFI ROHINGYA LETTER A	Aa	AA	AA
𐴁	HANIFI ROHINGYA LETTER BA	Ba	BA	BA
𐴂	HANIFI ROHINGYA LETTER PA	Pa	PA	PA
𐴃	HANIFI ROHINGYA LETTER TA	Ta	TA	THA
𐴄	HANIFI ROHINGYA LETTER TTA	Tda	TA	TA
𐴅	HANIFI ROHINGYA LETTER JA	Ja	JA	JA
𐴆	HANIFI ROHINGYA LETTER CA	Cha	CHA	CHA
𐴇	HANIFI ROHINGYA LETTER HA	Ha	HA	HA
𐴈	HANIFI ROHINGYA LETTER KHA	Kha	KHA	KHA
𐴉	HANIFI ROHINGYA LETTER FA	Pha	FA	FA
𐴊	HANIFI ROHINGYA LETTER DA	Da	THA	DHA
𐴋	HANIFI ROHINGYA LETTER DDA	Dha	DA	DA
𐴌	HANIFI ROHINGYA LETTER RA	Ra	RA	RA
𐴍	HANIFI ROHINGYA LETTER RRA	Rda	R̄A	RDA
𐴎	HANIFI ROHINGYA LETTER ZA	Za	ZA	ZA
𐴏	HANIFI ROHINGYA LETTER SA	Sa	SA	SA
𐴐	HANIFI ROHINGYA LETTER SHA	Sha	SHA	SHA
𐴑	HANIFI ROHINGYA LETTER KA	Ka	KA	KA
𐴒	HANIFI ROHINGYA LETTER GA	Ga	GA	GHA
𐴓	HANIFI ROHINGYA LETTER LA	La	LA	LA
𐴔	HANIFI ROHINGYA LETTER MA	Ma	MA	MA
𐴕	HANIFI ROHINGYA LETTER NA	Na	NA	NA
𐴖	HANIFI ROHINGYA LETTER WA	Wa	WA	WA
𐴗	HANIFI ROHINGYA LETTER SMALL WA	Kinnawa	KINNA WA	KINNAWA
𐴘	HANIFI ROHINGYA LETTER YA	Ya	YA	YA
𐴙	HANIFI ROHINGYA LETTER SMALL YA	Kinnaya	KINNA YA	KINNAYA
𐴚	HANIFI ROHINGYA LETTER NGA	Nga	GAN	GHAN
𐴛	HANIFI ROHINGYA LETTER NYA	Naiya	NAYYA	NAYA

Table 3: Romanized names for Rohingya letters as given in various sources.

	Proposed character name	Fig. 4	Fig. 5	Fig. 6
𐴝	HANIFI ROHINGYA VOWEL MARK A	Aafor	AFOR	AFOR
𐴞	HANIFI ROHINGYA VOWEL MARK I	Efor	[I]FOR	EIFOR
𐴟	HANIFI ROHINGYA VOWEL MARK U	Ufor	UFOR	UOFOR
𐴠	HANIFI ROHINGYA VOWEL MARK E	Aefor	EFOR	EAFOR
𐴡	HANIFI ROHINGYA VOWEL MARK O	Ofor	OFOR	WOFOR
𐴢	HANIFI ROHINGYA MARK SAKIN	—	—	—
𐴣	HANIFI ROHINGYA MARK NA KHANNA	Nakhonna	NAKONNA	NAKHANNA
◌𐴤	HANIFI ROHINGYA SIGN HARBAHAY	Haarbai	HARBAY	HAR BAY
◌𐴥	HANIFI ROHINGYA SIGN TAHALA	Tela	TELA	TAHA LA
◌𐴦	HANIFI ROHINGYA SIGN TANA	Tana	TANA	TA NA
◌𐴧	HANIFI ROHINGYA SIGN TASSI	Tossi	TASSI	TASSI

Table 4: Romanized names for Rohingya marks as given in various sources.

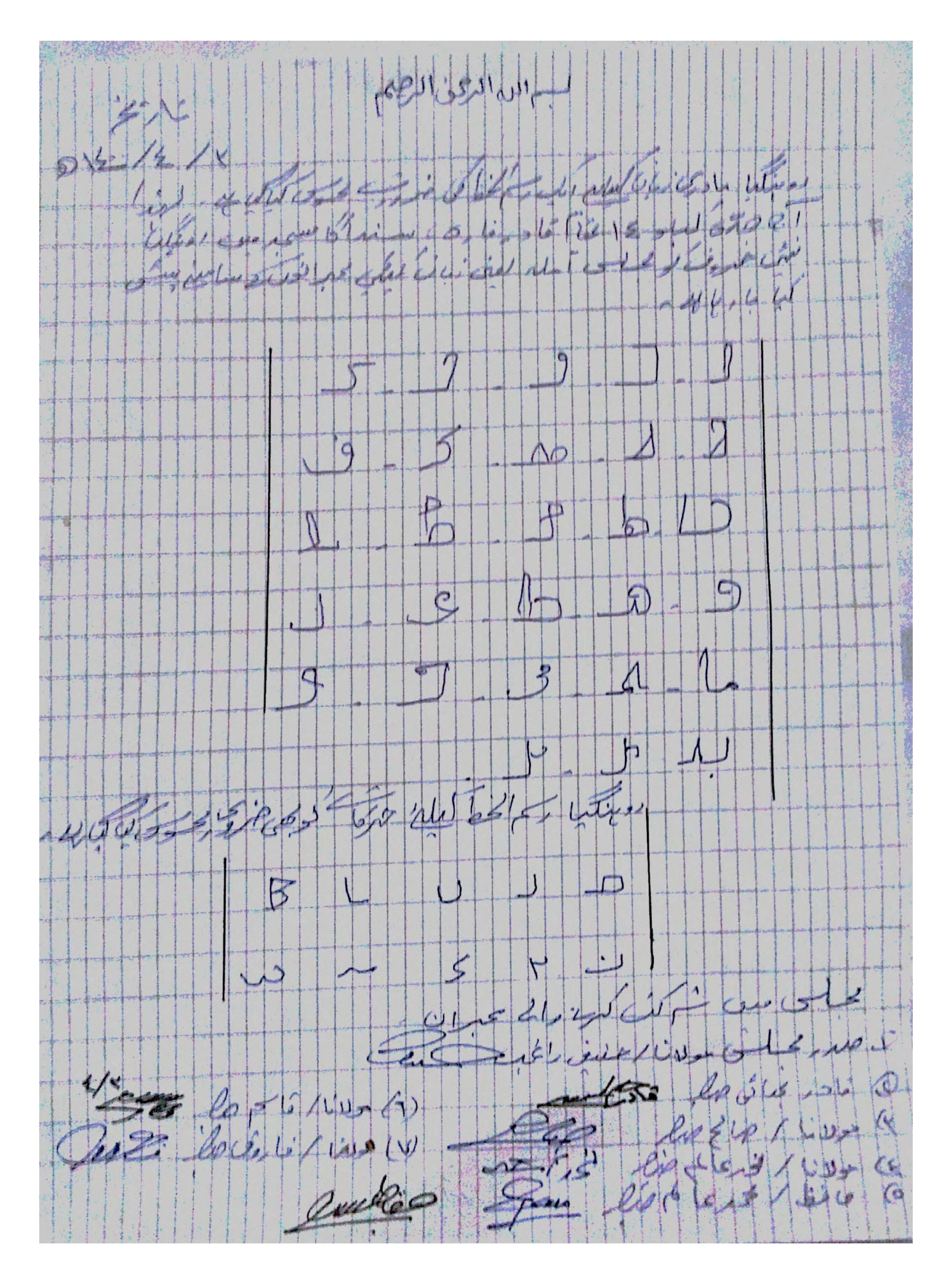

بسم الله الرحمن الرحيم

Figure 1: A document containing the Rohingya script as finalized by Hanif and others on February 19, 1980, with signatures of the individuals involved in the creation of the script.

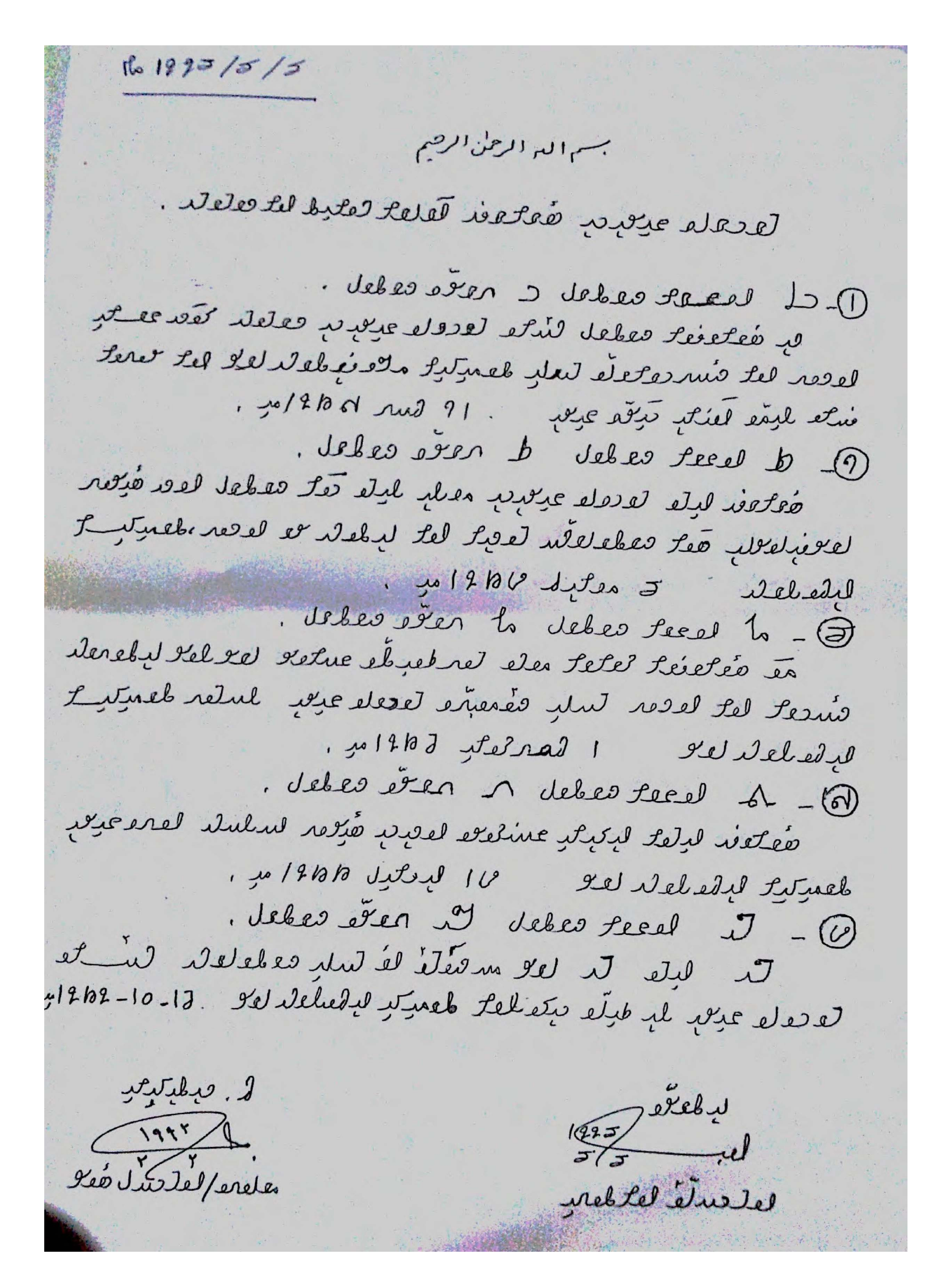


Figure 2: Description of some characters attached to the official chart shown in figure 1.



Figure 3: Chart of Rohingya script from a hand-written primer (from Ruwainggya Zuban Komiti (A): 1).

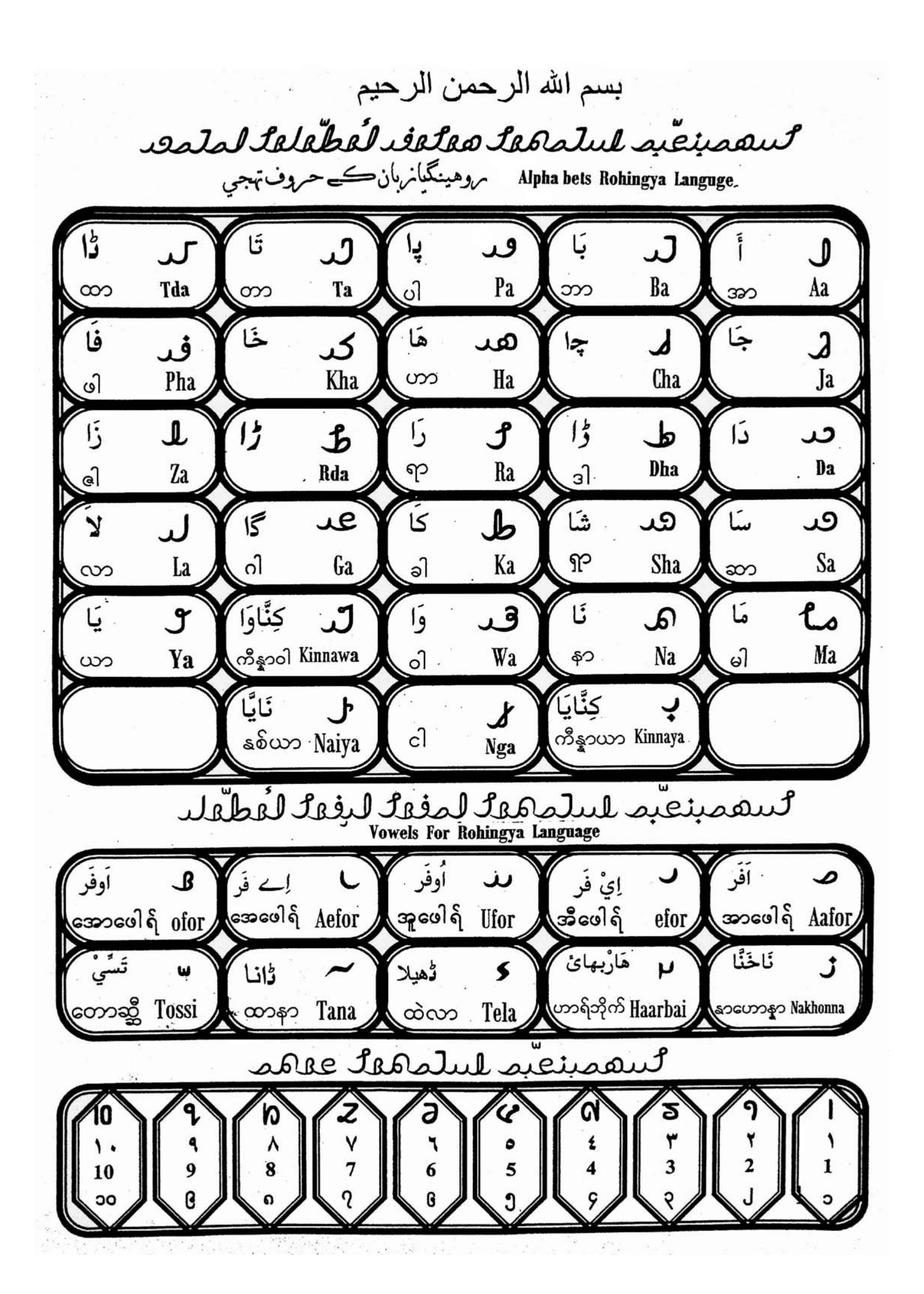


Figure 4: Chart showing Rohingya letters and signs with Urdu and Burmese names.

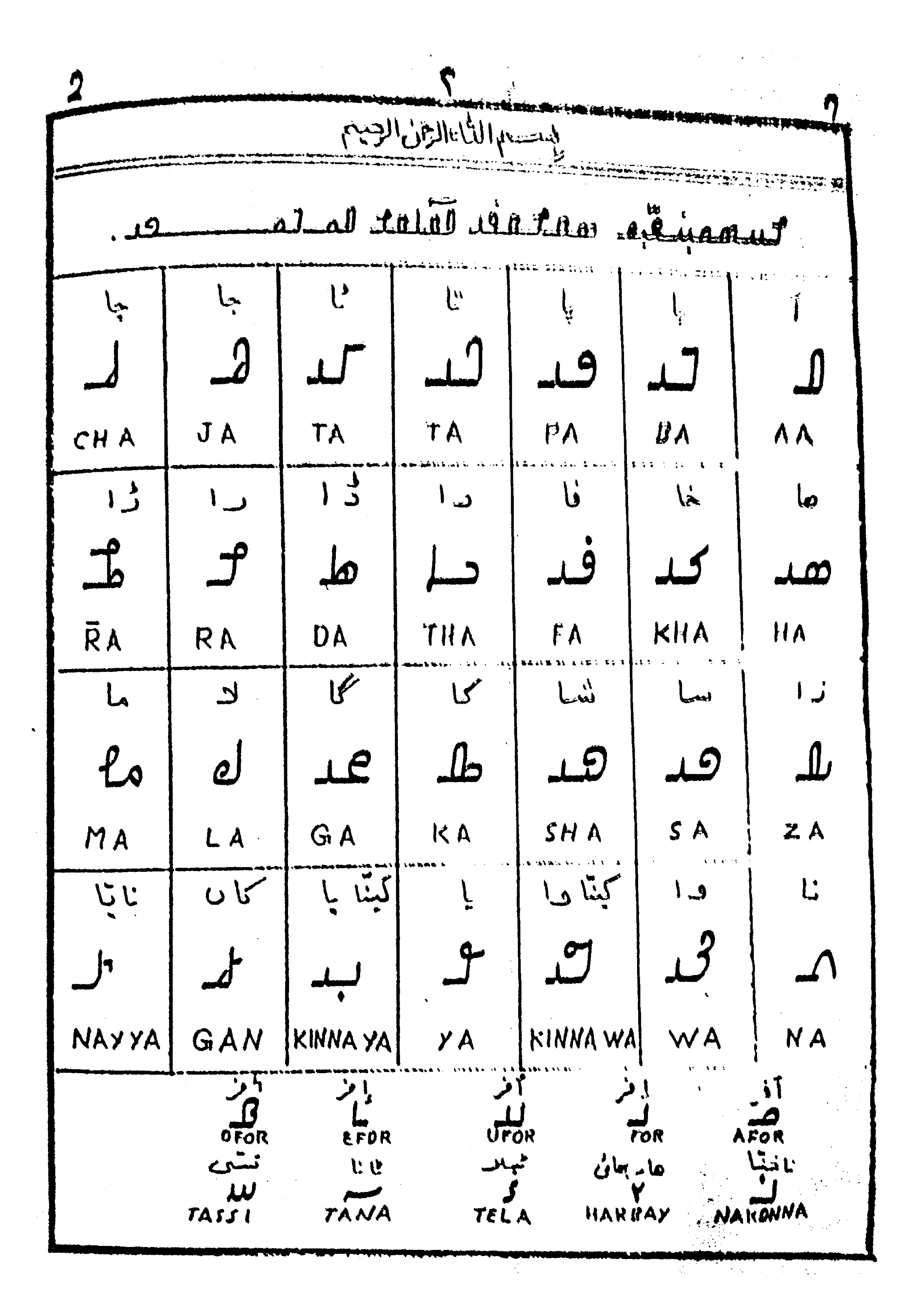


Figure 5: Chart showing Rohingya letters with Arabic correspondences and names in Latin transliteration.

بسم الله الرحمن الرحيم

TA	THA	PA	BA	AA
FA	KHA	HA	CHA	JA
ZA	RDA	RA	DA	DHA
LA	GHA	KA	SHA	SA
YA	KINNAWA	WA	NA	MA
		NAYA	GHAN	KINNAYA

WOFOR. EAFOR. UOFOR. EIFOR. AFOR.

TASSI . TA NA. TAHA LA. HAR BAY. NAKHANNA.

Figure 6: Chart showing Rohingya letters with Latin transliteration.

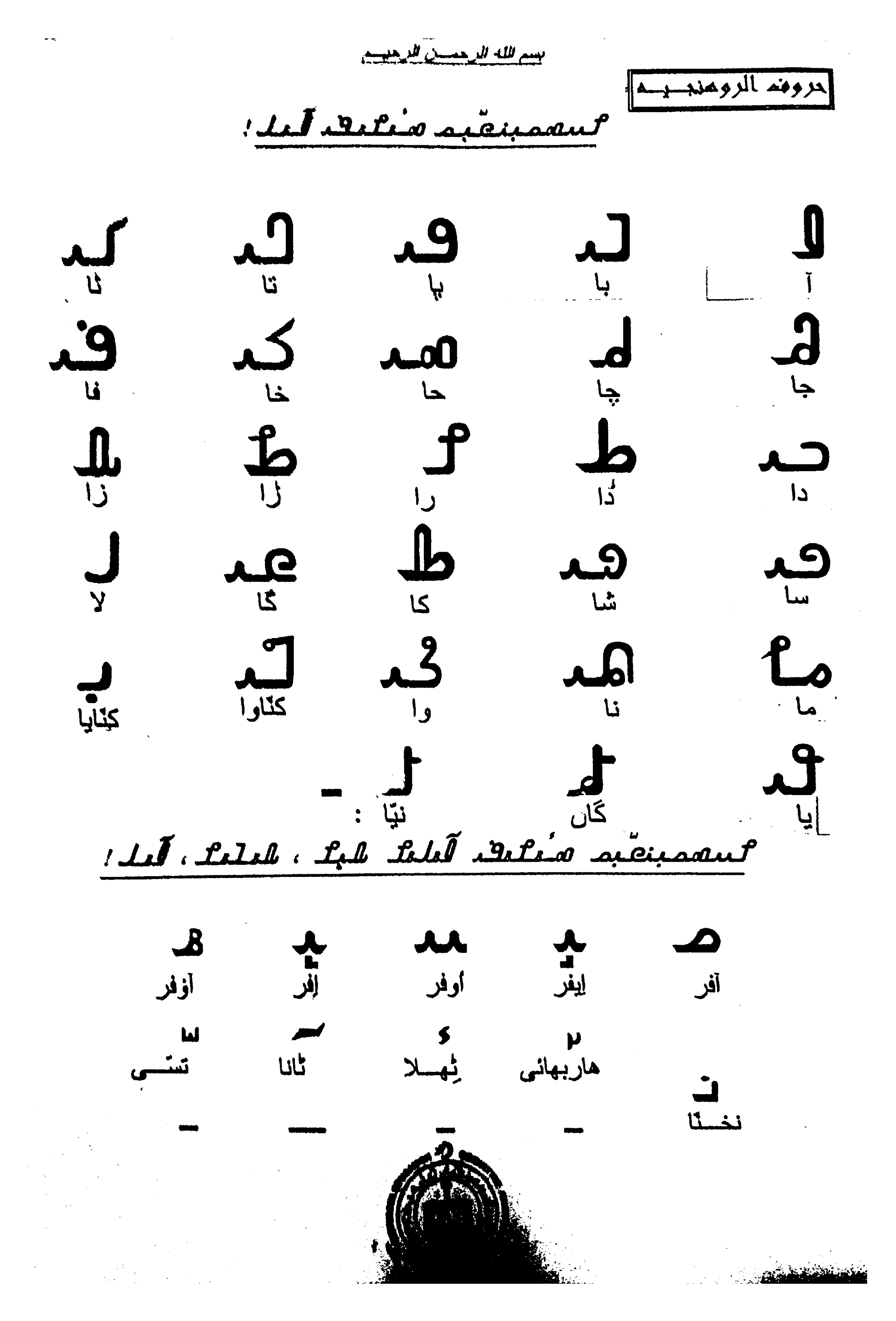


Figure 7: Page from a Rohingya primer showing Rohingya letters with names in Arabic transliteration.

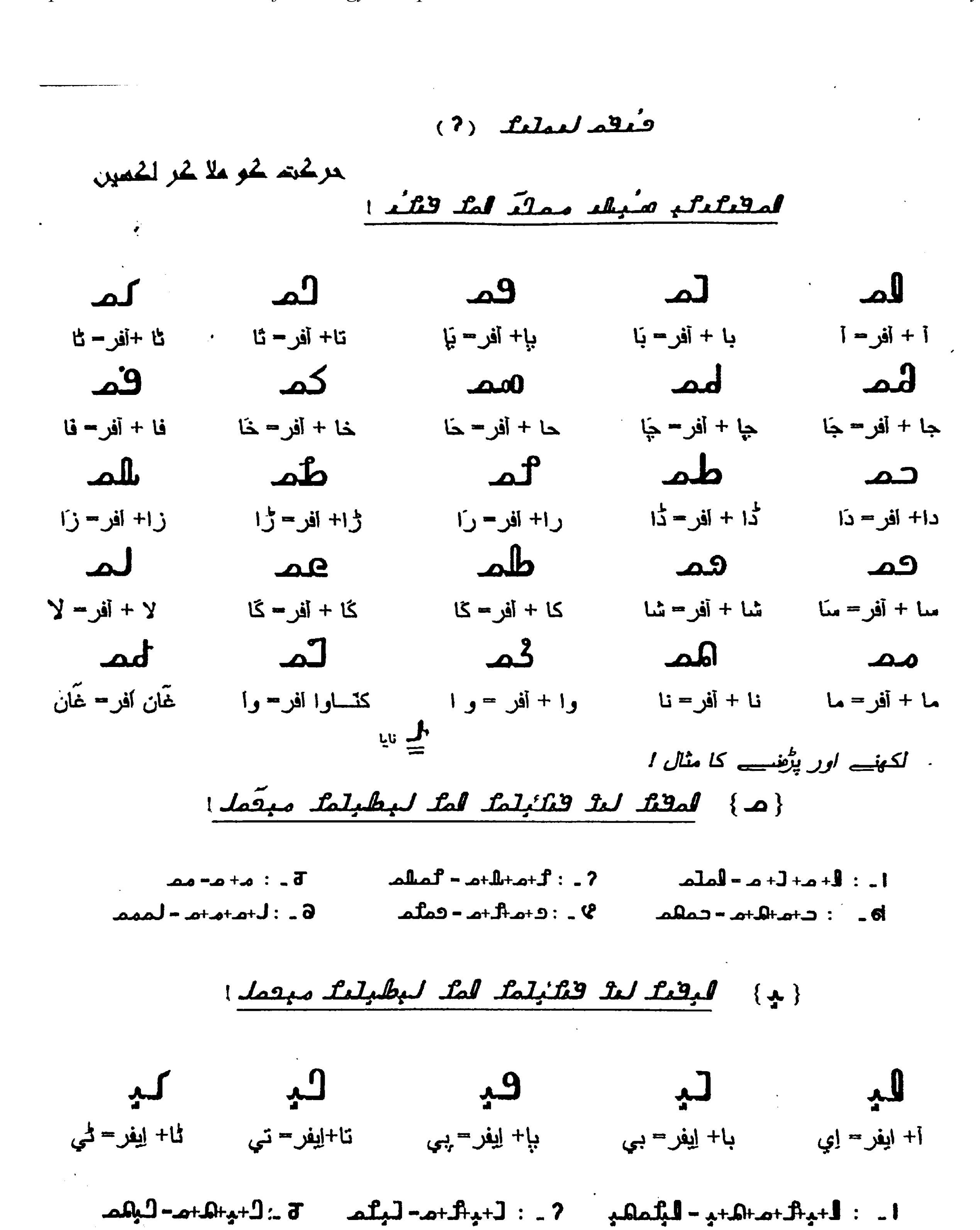


Figure 8: Page from a Rohingya primer showing the method of writing 𐴝 VOWEL MARK A and 𐴞 VOWEL MARK I.

[illegible]

[illegible]

آ+ اوفر= او　با+ اوفر= بُو　پا+ اُوفر= پُو　تا+ اوفر= تو　ٹا+ اوفر= ٹو

[illegible]

[illegible]

آ + اِ فر= ایے　با+ اِ فر= بے　پا + اِ فر= پے　تا + اِ فر= تے　ٹا + اِ فر= ٹے

[illegible]

[illegible]

آ+ آوفر= آوْ　با+ آوفر= بَو　پا+ آوفر= پَو　تا+ آوفر= تَو　ٹا+ آوفر= ٹَو

[illegible]

[illegible]

[illegible]

Figure 9: Page from a Rohingya primer showing the method of writing 𐴟 VOWEL MARK U, 𐴠 VOWEL MARK E, 𐴡 VOWEL MARK O.



Figure 10: Page from a Rohingya primer showing the method of writing 𐴣 LETTER NA KHANNA, ◌𐴤 TONE-1, ◌𐴥 TONE-2, ◌𐴦 TONE-3, ◌𐴧 SIGN TASSI.

Figure 11: Page from a Rohingya primer describing the use of SMALL WA and SMALL YA.





Figure 12: Page from a Rohingya primer showing the use of *tatweel*.

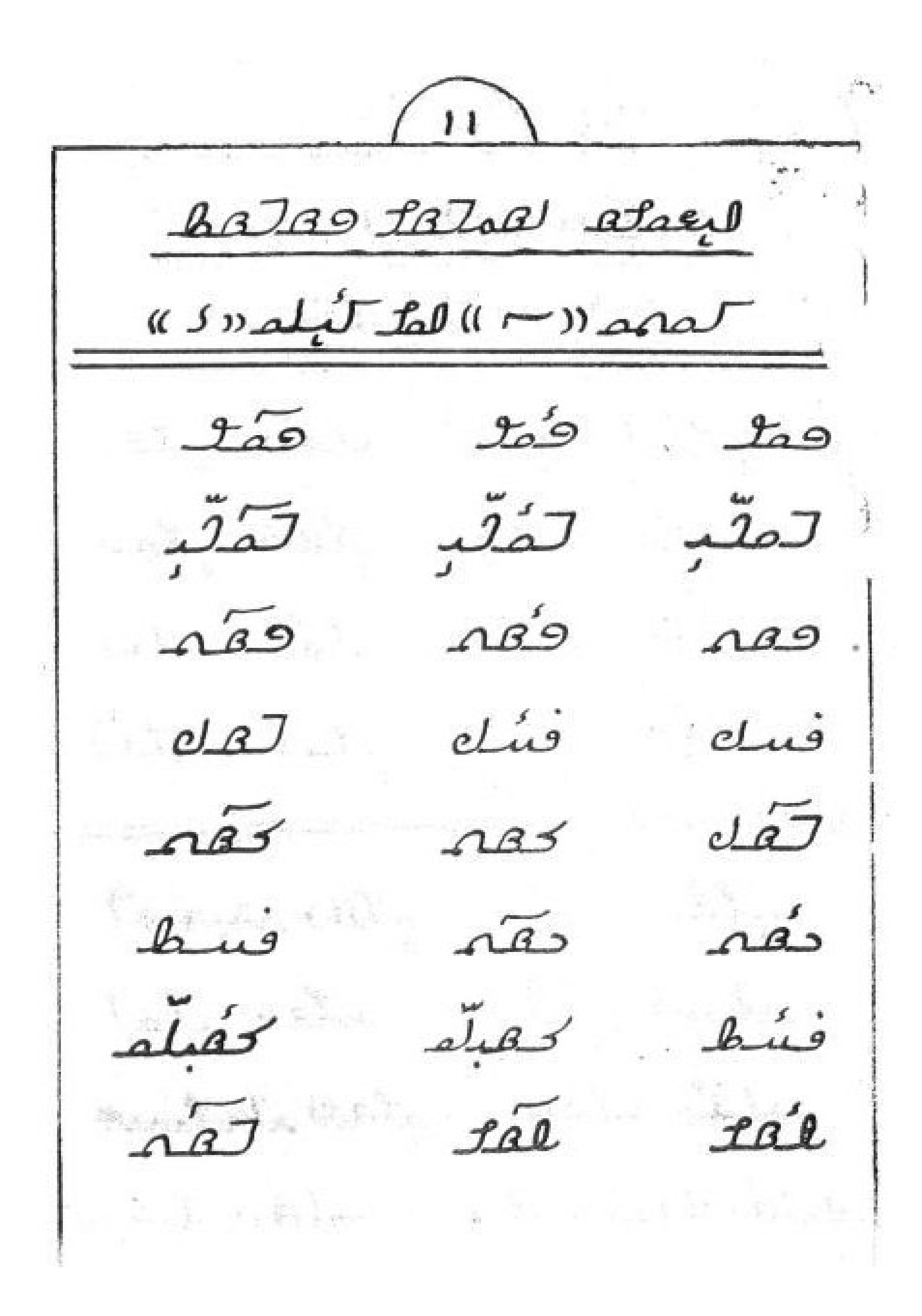


Figure 13: Table showing use of tonal signs from a hand-written primer (from Ruwainggya Zuban Komiti (A): 11).

Figure 14: Example of running Rohingya text from a hand-written primer (from Ruwainggya Zuban Komiti (B): 1).

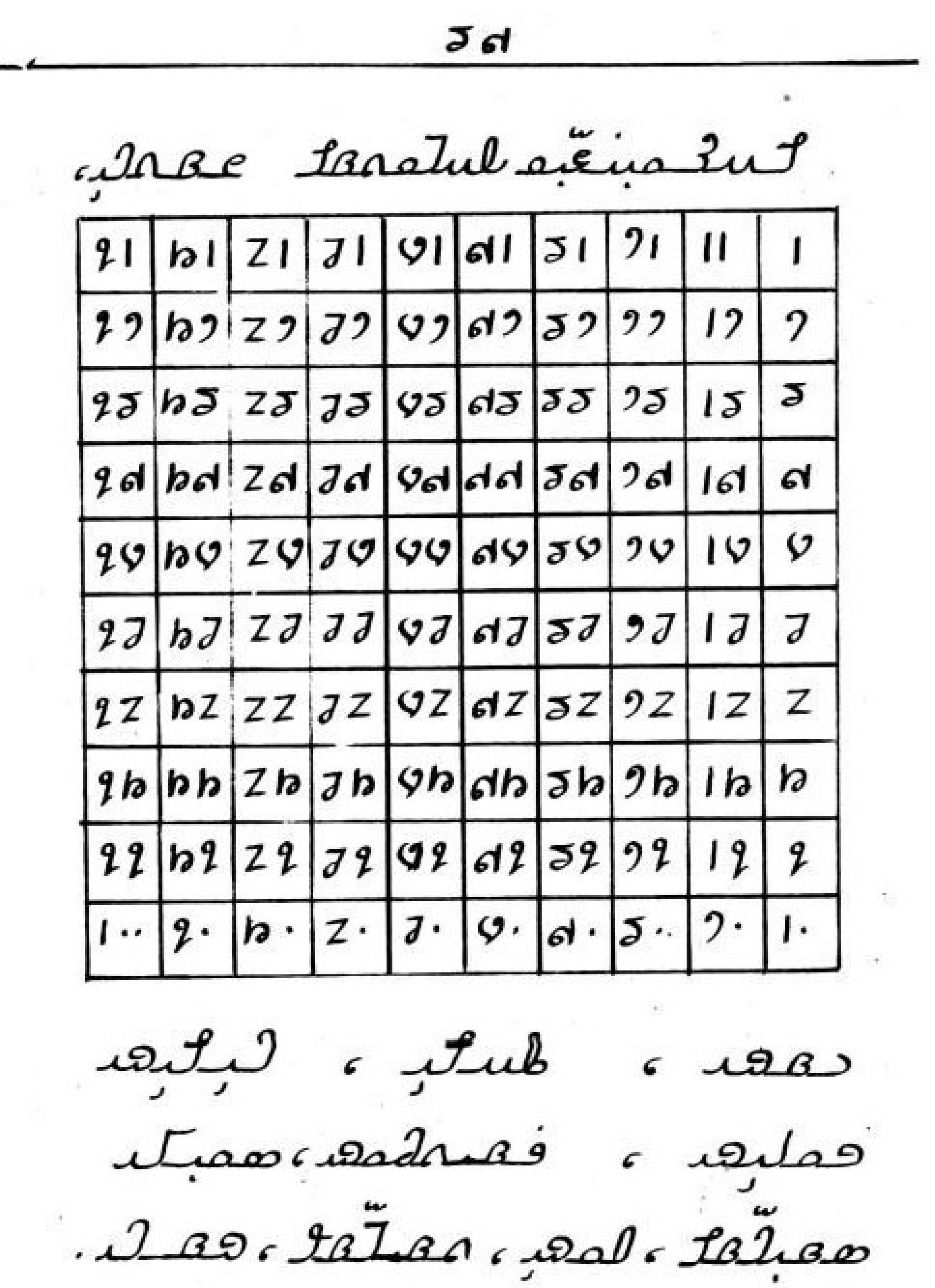


Figure 15: Chart of digits from a hand-written primer (from Ruwainggya Zuban Komiti (B): 34).

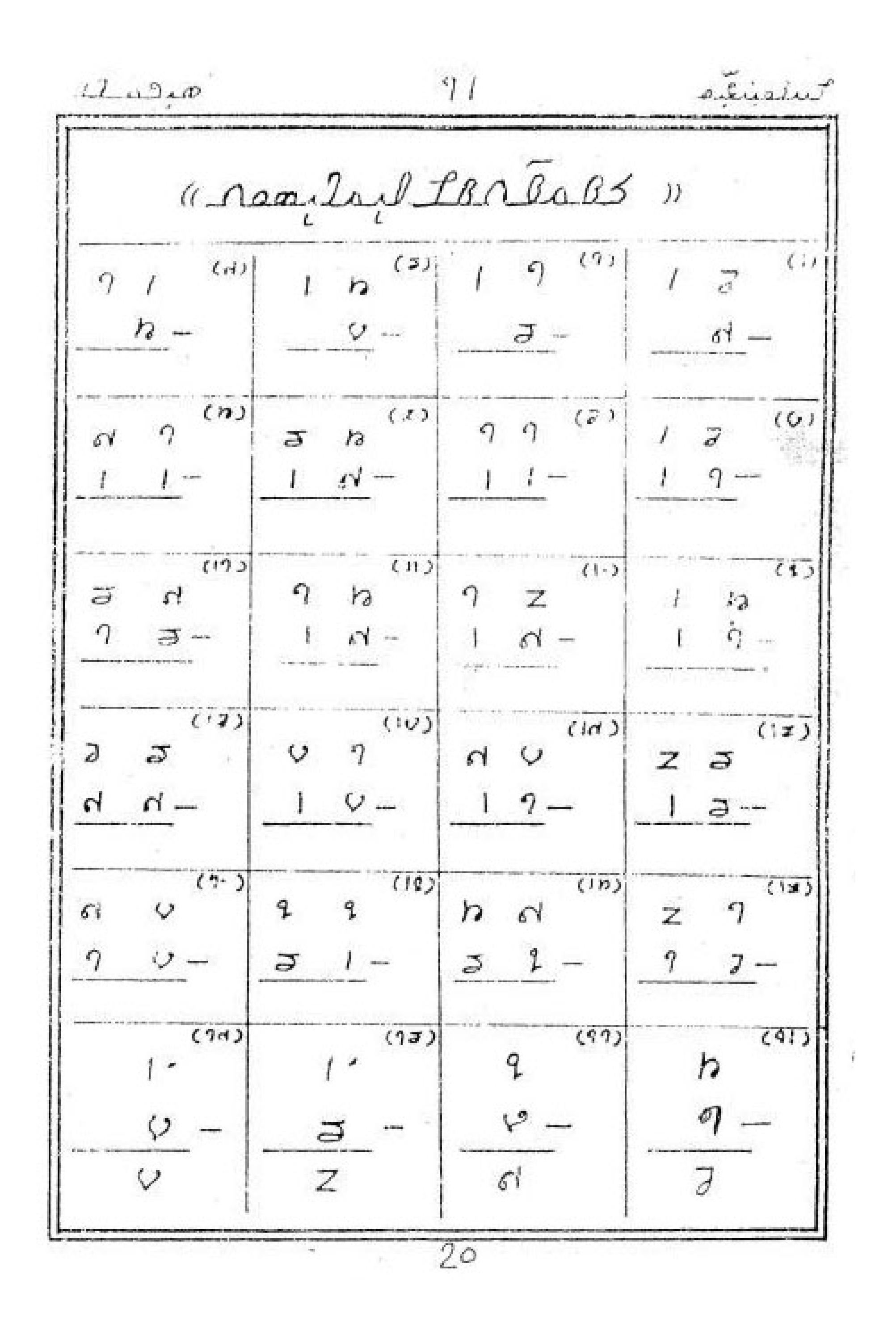


Figure 16: Excerpt from an primary-level arithmetic book written in Rohingya (from Ruwainggya Education Board Myanmar: 21).



قل جاء الحق وزهق الباطل ان الباطل كان زهوقا

5. December 2002.

The Rohaingya Language HAQ — DAR Monthly Founder Muhammad Iskander Imam

Figure 17: The cover page of *Haq-Dar*, a Rohingya language news weekly (December 5, 2002).



Figure 18: The cover page of *Serak*, a Rohingya language news weekly (October 13, 2007).

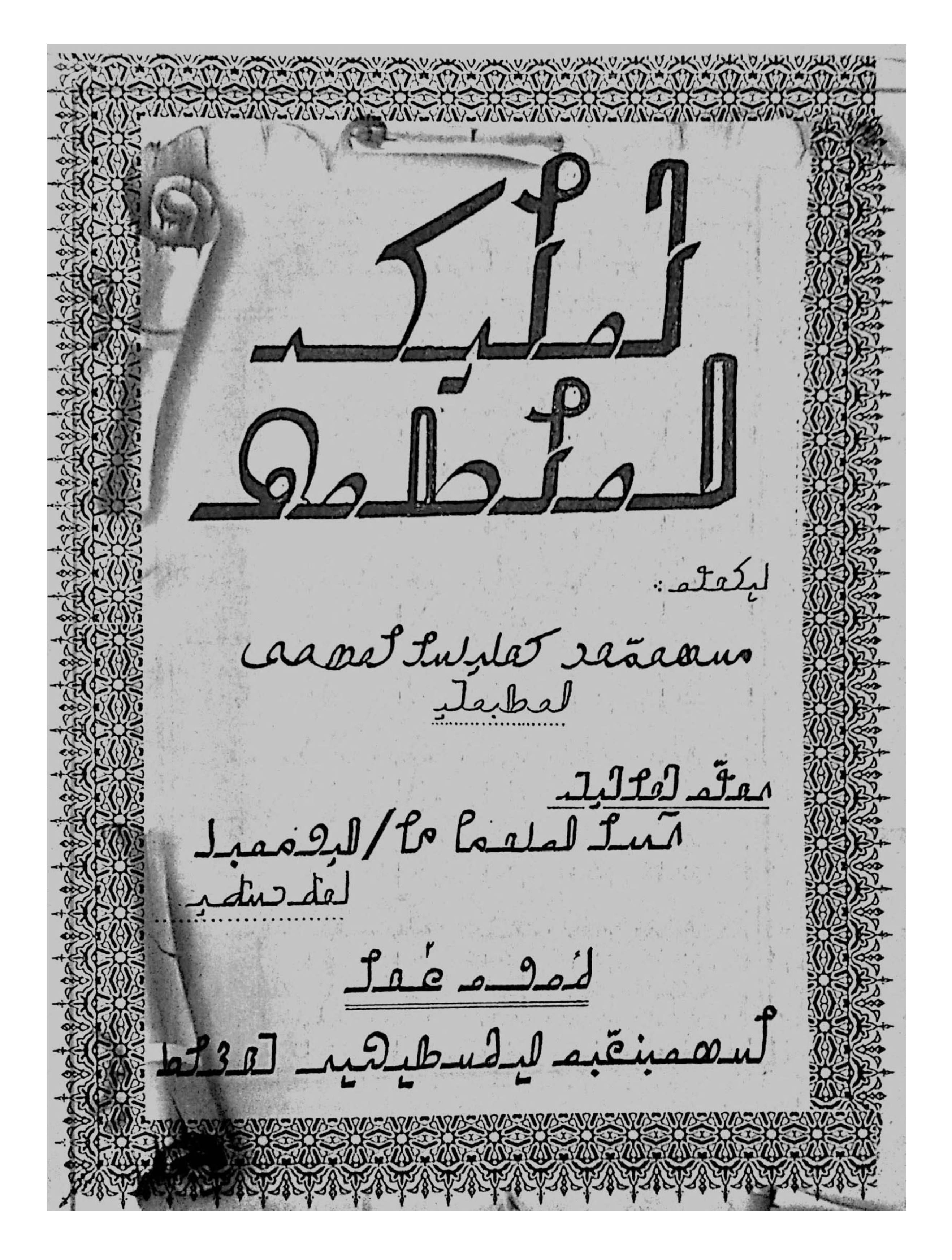


Figure 19: Cover page of *History of Arakan*. The vertical breaks in the letters 𐴃 TA and 𐴌 RA in the title word *tarikh arakan* are stylistic.

(تاریخ ارکان)

Figure 20: First page of *History of Arakan*.

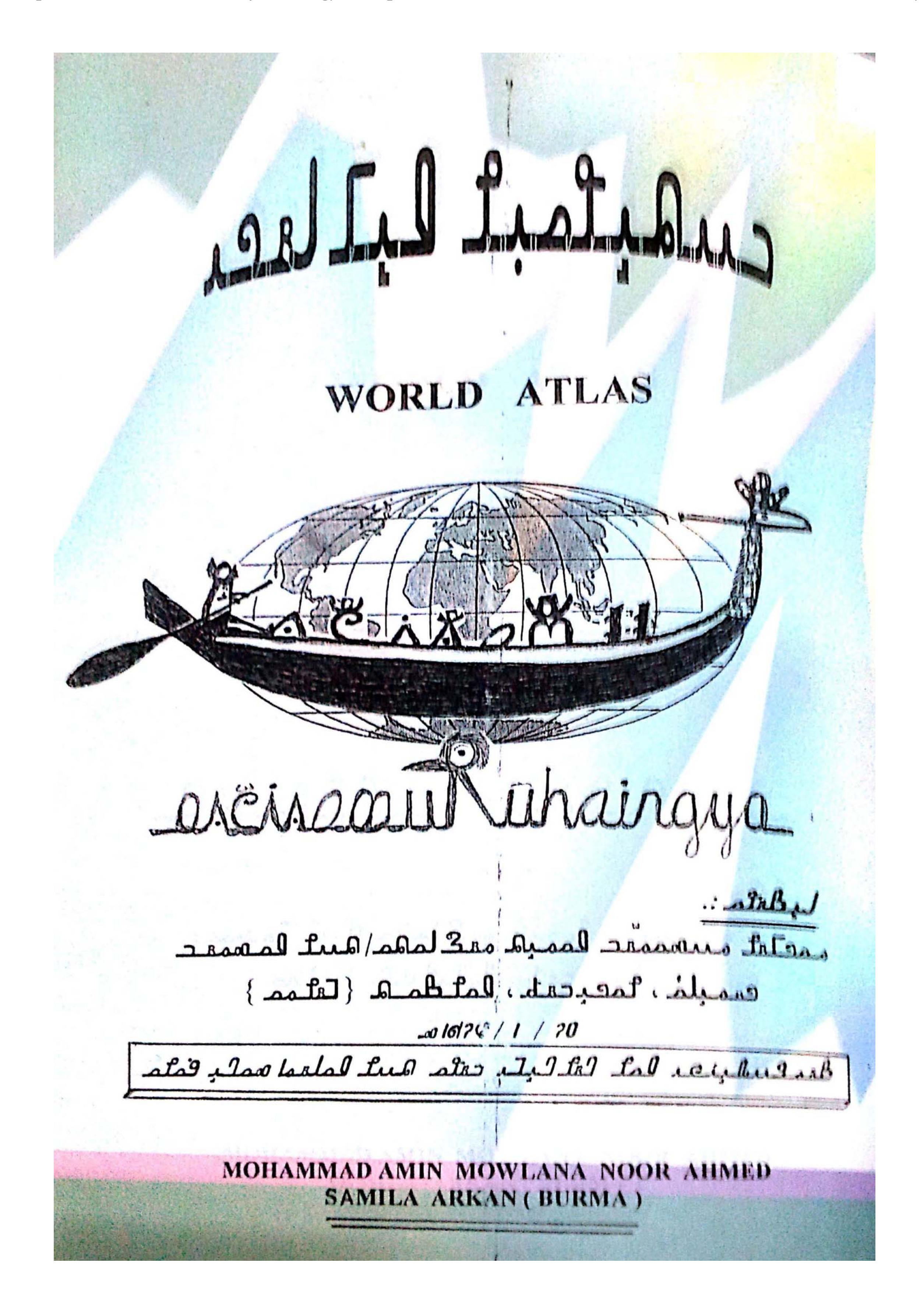


Figure 21: Cover of *World Atlas*.

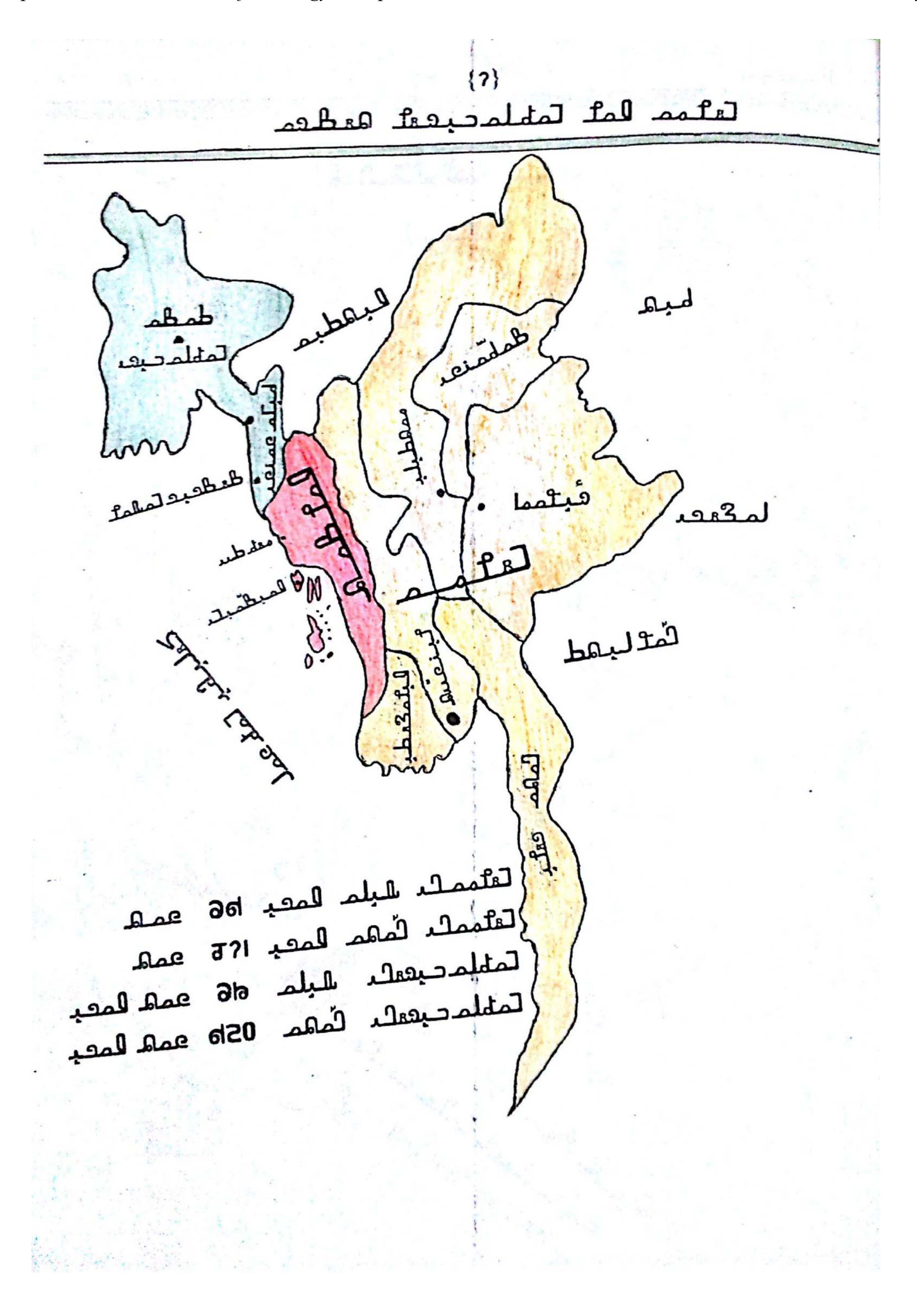


Figure 22: Page from *World Atlas* showing Myanmar (with the Arakan region highlighted) and Bangladesh.



Figure 23: Use of a Rohingya typeface in a digital video (from noorismail52 2011a: frame 3). The first character shown is 𐴱 hanifi rohingya digit one.



Figure 24: Use of a Rohingya typeface in a digital video (from noorismail52 2011b: frame 103). The first character shown is the number 𐴱𐴱 ‘11’ <𐴱 hanifi rohingya digit one, 𐴱 hanifi rohingya digit one>.

	G	K		G	K		G	K
A	𐴀	𐴀	KA	𐴑	𐴑	NA KHANNA	𐴣	𐴣
BA	𐴁	𐴁	GA	𐴒	𐴒	HARBAHAY	◌𐴤	◌𐴤
PA	𐴂	𐴂	LA	𐴓	𐴓	TAHALA	◌𐴥	◌𐴥
TA	𐴃	𐴃	MA	𐴔	𐴔	TANA	◌𐴦	◌𐴦
TTA	𐴄	𐴄	NA	𐴕	𐴕	TASSI	◌𐴧	◌𐴧
JA	𐴅	𐴅	WA	𐴖	𐴖	ZERO	𐴰	𐴰
CA	𐴆	𐴆	SMALL WA	𐴗	𐴗	ONE	𐴱	𐴱
HA	𐴇	𐴇	YA	𐴘	𐴘	TWO	𐴲	𐴲
KHA	𐴈	𐴈	SMALL YA	𐴙	𐴙	THREE	𐴳	𐴳
FA	𐴉	𐴉	NGA	𐴚	𐴚	FOUR	𐴴	𐴴
DA	𐴊	𐴊	NYA	𐴛	𐴛	FIVE	𐴵	𐴵
DDA	𐴋	𐴋	SAKIN	𐴢	𐴢	SIX	𐴶	𐴶
RA	𐴌	𐴌	AA	𐴝	𐴝	SEVEN	𐴷	𐴷
RRA	𐴍	𐴍	I	𐴞	𐴞	EIGHT	𐴸	𐴸
ZA	𐴎	𐴎	U	𐴟	𐴟	NINE	𐴹	𐴹
SA	𐴏	𐴏	E	𐴠	𐴠			
SHA	𐴐	𐴐	O	𐴡	𐴡			

Table 5: Comparison of digitized Rohingya typefaces: “Rohingya Gonya Leyka Noories” (‘G’) and “Rohingya Kuna Leyka Noories” (‘K’) designed by Muhammad Noor.